



www.volsu.ru

DOI: <https://doi.org/10.15688/mpcm.jvolsu.2021.3.4>

УДК 004.4.068

ББК 22.182

Дата поступления статьи: 09.06.2021

Дата принятия статьи: 30.07.2021



ПРОГНОЗИРОВАНИЕ РИСКОВ ОРГАНИЗАЦИИ, ЭКСПЛУАТИРУЮЩЕЙ ТРАНСПОРТНЫЕ СРЕДСТВА НА ПРИРОДНОМ ГАЗЕ, С ИСПОЛЬЗОВАНИЕМ СКОРИНГ-МОДЕЛИ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ПРИ НАЛИЧИИ ЭКСПЕРТНЫХ ОГРАНИЧЕНИЙ

Андрей Александрович Евстифеев

Кандидат технических наук,
доцент кафедры № 12 компьютерных систем и технологий,
Национальный исследовательский ядерный университет,
Московский инженерно-физический институт
aaevstifeev@mephi.ru
<https://orcid.org/0000-0003-3354-0487>
Каширское шоссе, 31, 115409 г. Москва, Российская Федерация

Аннотация. В работе предложен способ и описана математическая модель экспресс-анализа привлекательности эксплуатации транспортных средств на природном газе для автотранспортного предприятия. Предложенное решение базируется на скоринг-модели логистической регрессии, используемой банками для оценки кредитоспособности заемщика. Для повышения качества результатов модель расширена набором экспертных ограничений, сформулированных в виде правил. В процессе анализа выявлены признаки, требующие квантования, поскольку отдельные интервалы значений оказались по-разному связаны с риском. Разработанная математическая модель реализована в виде программного обеспечения на языке программирования высокого уровня, информация модели хранится в системе управления базами данных и интегрирована с информационной системой поддержки принятия управленческих решений при эксплуатации транспортных средств на природном газе. Проведена проверка разработанной математической модели на тестовой обучающей выборке. Результаты тестирования показали удовлетворительную точность предложенной модели на уровне 77 % без использования экспертных ограничений и 79 % с их использованием. При этом доля ошибок второго

рода составила 2,7 %, а ошибок первого рода — 7,2 %, что говорит о том, что модель достаточно консервативна, и относительно высокая доля соответствующих требованиям транспортных средств получила отказ.

Ключевые слова: скоринг-модель, логистическая регрессия, эксплуатация транспортных средств, принятие решений, сложная техническая система.

Введение

Современное общество находится в постоянном поиске альтернативных возобновляемых источников энергии, в том числе новых видов моторного топлива. Основными предпосылками к переходу автотранспортных предприятий и частных владельцев транспортных средств на альтернативное топливо являются: государственная фискальная политика в области повышения экологических требований к транспортным средствам, снижение доходов населения и высокая стоимость традиционных нефтяных топлив. Одним из наиболее распространенных альтернативных видов моторного топлива является природный газ [6].

Первые промышленные партии транспортных средств с газобаллонным оборудованием для сжатого природного газа были произведены отечественными автомобильными заводами в конце 40-х — начале 50-х гг. XX в. [2]. В период с начала 80-х до середины 90-х гг. XX в. была построена базовая сеть автомобильных газонаполнительных заправочных станций (АГНКС) и накоплен опыт эксплуатации транспортных средств на газомоторном топливе [5; 16].

В настоящее время на территории Российской Федерации эксплуатируется около 200 тыс. автомобилей на природном газе. Парк газобаллонных транспортных средств неоднороден. Присутствуют автомобили как заводского изготовления, так и переоборудованные, в том числе и незаконно [6]. На итоговый производственный результат от использования транспортного средства влияет множество факторов: пробег, участие в авариях и ДТП, качество и своевременность технического обслуживания, количество водителей, вид собственности (полная или лизинг) и целый ряд других.

Одной из задач организации, эксплуатирующей транспортные средства, является оперативное принятие решений о рациональности приобретения и продолжения эксплуатации транспортных средств [17]. Анализ подобного рода информации для сложных технических систем [4; 13], к которым относится современный автомобиль, рекомендуется выполнять с использованием экспертных систем [15]. Однако можно сократить затраты времени, сведя задачу анализа данного вида к задаче бинарной классификации, одним из вариантов реализации которой является кредитный скоринг, используемый банками при анализе кредитоспособности потенциального заемщика [1; 3; 7]. Одним из вариантов реализации скоринг-модели является разработка программной реализации алгоритма на языке Python [10; 12; 14], с организацией хранения информации в системе управления базами данных [9] с поддержкой языка программирования высокого уровня [11].

1. Скоринг-модель логистической регрессии при наличии экспертных ограничений

1.1. Математическая постановка задачи

Задача кредитного скоринга является задачей бинарной классификации.

Пусть Ω — множество производимых промышленностью транспортных средств с газобаллонным оборудованием, которые могут быть приобретены эксплуатирующей организацией. Каждому потенциальному транспортному средству $\omega \in \Omega$ ставится в соответствие его признаковое описание $x = (x_1(\omega), \dots, x_n(\omega)) \in X = R^n$, а также значение зависимой переменной $Y = \{0; 1\}$. Зависимая переменная описывает класс транспортного средства, метка 1 соответствует классу «хорошее», а метка 0 — классу «плохое» транспортное средство, из-за которых эксплуатирующая организация понесет убытки.

В задаче скоринга требуется построить такую модель, которая по признаковому описанию x потенциального транспортного средства будет предсказывать значение зависимой переменной y . Предполагаем, что на множестве $A = X \times Y$ существует некоторое вероятностное распределение с плотностью $p(x, y)$. Из множества A выбирается выборка $\Omega^* = \{(x(\omega_i), y(\omega_i))\}_{i=1}^m$ — это набор случайно и независимо выбранных наблюдений, то есть описания транспортных средств. Полученная выборка Ω^* называется обучающей выборкой, на ее основе строится модель.

Для решения поставленной задачи обычно используется принцип максимума апостериорной вероятности. Апостериорная вероятность того, что транспортное средство принадлежит «хорошему» или «плохому» классу, может быть посчитана с использованием формулы Байеса:

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{p(x|y)p(y)}{p(x)}. \quad (1)$$

При решении задачи бинарной классификации условное распределение зависимой переменной представляет собой распределение Бернулли [3]:

$$p(y|\mathbf{x}, \omega) = \text{Ber}(y | \text{sigm}(\omega^T x)), \quad (2)$$

где $\text{sigm}(\omega^T x) = \frac{1}{1 + \exp^{-\omega^T x}}$ — логистическая функция (сигмоида), а $\omega^T x = \sum_{i=1}^n \omega_i x_i = = 1$ — линейная разделяющая гиперплоскость. Тогда вероятности того, что транспортное средство принадлежит к классам «плохих» и «хороших» равны соответственно:

$$P(y = +1|\mathbf{x}, \omega) = \text{sigm}(\omega^T x); \quad (3)$$

$$P(y = -1|\mathbf{x}, \omega) = 1 - \text{sigm}(\omega^T x). \quad (4)$$

Выражения (3) и (4) можно представить в общем виде:

$$p(y|\mathbf{x}, \omega) = \text{sigm}(y\omega^T x) = \frac{1}{1 + \exp^{-y\omega^T x}}. \quad (5)$$

Поскольку наблюдения в обучающей выборке независимы, поэтому функция правдоподобия будет выглядеть следующим образом:

$$L(\omega|\Omega^*) = \prod_{i=1}^m p(y_i|x_i, \omega) = \prod_{i=1}^m \text{sigm}(y_i \sum_{j=1}^n \omega_j x_{ij}). \quad (6)$$

Используя принцип максимума правдоподобия, получаем оценку вектора параметров [10]:

$$\hat{\omega}_{ML} = \arg \max_{\omega} \left\{ \prod_{i=1}^m \text{sigm} \left(y_i \sum_{j=1}^n \omega_j x_{ij} \right) \right\}. \quad (7)$$

Прологарифмируем функцию правдоподобия 7 и будем решать задачу минимизации:

$$\hat{\omega}_{ML} = \arg \min_{\omega} \left\{ - \sum_{i=1}^m \ln \left(\text{sigm} \left(y_i \sum_{j=1}^n \omega_j x_{ij} \right) \right) \right\}. \quad (8)$$

В соответствии с (8) оценивают коэффициенты в классической нерегуляризованной логистической регрессии. Предположим, что априорной плотностью распределения параметра является нормальное распределение с нулевым математическим ожиданием и дисперсией r . В модели дисперсия будет являться случайной величиной.

Тогда совместное распределение вектора параметров будет иметь вид:

$$p(\omega|r) \propto \prod_{i=1}^n \left[\left(\frac{1}{r_i} \right)^{0.5} \exp \left(- \sum_{i=1}^n \frac{\omega_i^2}{2r_i} \right) \right]. \quad (9)$$

Параметры с малым значением r_i могут быть удалены из модели, а остальные параметры будут называться релевантными. Здесь r_i — является гиперпараметром модели. Попробуем величины, обратные дисперсиям, использовать в качестве штрафных функций. Тогда предполагаем, что априорная плотность распределения величин, обратных дисперсиям, является гамма-распределением:

$$p\left(\frac{1}{r} | \alpha, \beta\right) \propto \prod_{i=1}^n \left[\left(\frac{1}{r_i} \right)^{\alpha-1} \exp \left(- \frac{\beta}{r_i} \right) \right]. \quad (10)$$

Из (10) видно, что обратная дисперсия зависит от двух параметров гамма-распределения α, β . Для облегчения процесса подбора параметров предположим, что они являются функциями от одного и того же параметра μ .

Требования. Для случайной величины, имеющей гамма-распределение [10], известно, что математическое ожидание $E\left(\frac{1}{r_i}\right) = \frac{\alpha}{\beta}$ и дисперсия $Var\left(\frac{1}{r_i}\right) = \frac{\alpha}{\beta^2}$.

Рассмотрим отношение $\frac{\sqrt{Var(1/r_i)}}{E(1/r_i)} = \frac{1}{\sqrt{\alpha}}$: если $\frac{\sqrt{Var(1/r_i)}}{E(1/r_i)} = \frac{1}{\sqrt{\alpha}} \rightarrow 0$, то значит все распределения дисперсий r_i сконцентрированы возле математического ожидания, тогда можно сказать, что оцененные дисперсии практически фиксированы и равны единице при $\alpha \simeq \beta$. Если $\frac{\sqrt{Var(1/r_i)}}{E(1/r_i)} = \frac{1}{\sqrt{\alpha}} \rightarrow 1$, то априорные распределения практически равномерны. При $r_i \rightarrow 0 : \ln r_i \rightarrow \infty$ критерию выгодно уменьшать все дисперсии. Но в этом случае невозможно выполнить ограничения, предписывающие достаточно хорошо приближать обучающую совокупность. Из-за этого противоречия критерий проявляет ярко выраженную склонность к чрезмерной селективности отбора признаков, подавляя большинство из них, в том числе и релевантные.

Одним из наборов функций, удовлетворяющих требованиям, является:

$$\alpha = 1 + \frac{1}{2\mu} \beta = \frac{1}{2\mu}. \quad (11)$$

С учетом всех предположений об априорных распределениях вектора параметров и гиперпараметров получаем следующую оценку вектора параметров [10]:

$$\hat{\omega}_{MAP} = \arg \underbrace{\max}_{\omega} \{p(y|\omega, x)p(\omega|r)p(r|\alpha, \beta)\}; \quad (12)$$

$$\hat{\omega}_{MAP} = \arg \underbrace{\min}_{\omega} \{-\ln p(y|\omega, x) - \ln p(\omega|r) - \ln p(r|\alpha, \beta)\}. \quad (13)$$

Из всего вышесказанного критерий обучения будет иметь вид:

$$J(\omega, r) = - \sum_{i=1}^m \ln \left(\text{sigm} \left(y_i \sum_{i=1}^n \omega_j x_{ij} \right) \right) + \quad (14)$$

$$+ \frac{1}{2} \sum_{i=1}^n \frac{\omega_i^2}{r_i} + \left(1 + \frac{1}{\mu} \right) \sum_{i=1}^n \ln r_i + \frac{1}{2\mu} \sum_{i=1}^n \ln r_i \rightarrow \min.$$

Критерий (14) является базовой моделью логистической регрессии.

1.2. Экспертные ограничения

При разработке модели для повышения качества результатов формируемой оценки были добавлены экспертные ограничения, основанные на эмпирических исследованиях. В частности, установлено, что риск возрастает с увеличением общей стоимости ремонта. Данный признак вошел в модель с положительным коэффициентом и было введено следующее ограничение:

$$\omega_{required_{limit}} \geq 0. \quad (15)$$

Обратная ситуация со сроком ремонта. Чем меньше необходимый срок, тем больше риски, и ограничение должно быть следующим:

$$\omega_{required_{term}} < 0. \quad (16)$$

Также логично ввести наличие предоплаты по запасным частям и ремонтным работам с положительным коэффициентом:

$$\omega_{required_{provision}} \geq 0. \quad (17)$$

В процессе анализа были выявлены признаки, требующие квантования, то есть разбиения на интервалы значений, где интервалы по-разному связаны с риском. Например, если в $(i+1)$ -м интервале вероятность получения убытка или факта отказа оплаты больше, чем в i -м, то было введено следующее ограничение:

$$\omega_{i+1} \geq \omega_i. \quad (18)$$

После дополнения модели логистической регрессии экспертными ограничениями была сформирована задача минимизации

$$\underbrace{\min}_{\omega \in R^n} \left(- \sum_{i=1}^m \ln \left(\text{sigm} \left(y_i \sum_{i=1}^n \omega_j x_{ij} \right) \right) + \right. \\ \left. + \frac{1}{2} \sum_{i=1}^n \frac{\omega_i^2}{r_i} + \left(1 + \frac{1}{\mu} \right) \sum_{i=1}^n \ln r_i + \frac{1}{2\mu} \sum_{i=1}^n \ln r_i \right), \quad (19)$$

при линейных ограничениях $\varphi(\omega) \leq 0$, где $\varphi(\omega)$ — выпуклая функция.

1.3. Решение задачи минимизации

Для решения поставленной задачи предлагается использовать метод штрафных функций. Рассмотрим задачу минимизации с ограничениями [10]:

$$f^* = \underbrace{\min}_{\omega \in R^n} F(x), \quad (20)$$

$$\varphi_i(x) \leq 0, \text{ при } i = 1, \dots, m, \quad (21)$$

$$\varphi_i(x) = 0, \text{ при } i = m + 1, \dots, k. \quad (22)$$

Пусть ограничения задают множество $G \in R^n$. Определим индикаторную функцию множества G следующим образом [8]:

$$\delta(x|G) = \begin{cases} 0, & x \in G, \\ +\infty, & x \notin G. \end{cases} \quad (23)$$

Допустим: $F(x) = f(x) + \delta(x|G)$. Тогда задача (20) эквивалентна следующей задаче безусловной минимизации:

$$\underbrace{\min}_{\omega \in R^n} F(x). \quad (24)$$

Пусть $\delta(x|G) = \lim_{k \rightarrow \infty} \delta_k(x|G)$. $\delta_k(x|G)$ назовем штрафными функциями и вместо задачи (24) будем решать следующую задачу [18]:

$$\underbrace{\min}_{\omega \in R^n} F_k(x) = \underbrace{\min}_{\omega \in R^n} f(x) + \delta_k(x|G), \quad (25)$$

$$x_k^* = \arg \underbrace{\min}_{\omega \in R^n} F_k(x). \quad (26)$$

А задача минимизации примет вид:

$$\underbrace{\min}_{\omega \in R^n} \left(- \sum_{i=1}^m \ln \left(\text{sigm} \left(y_i \sum_{i=1}^n \omega_j x_{ij} \right) \right) + \right. \\ \left. + \frac{1}{2} \sum_{i=1}^n \frac{\omega_i^2}{r_i} + \left(1 + \frac{1}{\mu} \right) \sum_{i=1}^n \ln r_i + \frac{1}{2\mu} \sum_{i=1}^n \ln r_i + \delta_k(\omega) \right), \quad (27)$$

где $\delta_k(\omega) = c_k(\varphi(\omega))^2$, $c_{k+1} > c_k$, $k = 1, 2, \dots, n$.

Для минимизации выведенного критерия используем покоординатный спуск. Для этого найдем градиент функции (27) по ω :

$$\frac{\partial}{\partial \omega_k} J(\omega, r) = - \sum_{i=1}^m (1 - \text{sigm}(y_i \omega^T x_i)) y_i x_{ik} + \frac{\omega_k}{r_k} + 2c_k \varphi(\omega_k) \varphi'(\omega_k), \quad (28)$$

$$\begin{aligned} \frac{\partial}{\partial \omega_k \omega_i} J(\omega, r) = & - \sum_{i=1}^m \text{sigm}(y_i \omega^T x_i) (1 - \text{sigm}(y_i \omega^T x_i)) x_{il} x_{ik} + \\ & + \frac{1}{r_k} + 2c_k (\varphi'(\omega_k)^2) + \varphi''(\omega_k) \varphi(\omega_k), \end{aligned} \quad (29)$$

$$\frac{\partial}{\partial r_k \omega_i} J(\omega, r) = - \frac{1}{2} \frac{\omega_k^2}{r_k^2} + \frac{1}{2} \left(1 + \frac{1}{\mu} \right) \frac{1}{r_k} - \frac{1}{2\mu} \frac{1}{r^2}. \quad (30)$$

Приравняем производную по r к нулю и найдем значение, соответствующее минимальному значению целевой функции при фиксированном наборе весов:

$$r_k = \frac{\mu \omega_k^2 + 1}{\mu + 1}. \quad (31)$$

Множество параметров μ приводит к набору моделей, среди которых необходимо выбрать имеющую наилучшую обобщающую способность. Одним из хорошо себя зарекомендовавших методов является процедура кросс-валидации по блокам [7]. Основная идея заключается в последовательном изменении параметра регуляризации и использовании коэффициентов модели, найденных при предыдущем значении параметров в качестве начального приближения новой модели. С точки зрения реализации это, по сути, означает добавление внешнего цикла, где увеличивается параметр μ .

Финальным этапом разработки является ранжирование и перевод коэффициентов модели в скоринговые баллы. Скоринговый балл Sb можно рассчитать умножением оценки коэффициентов логистической регрессии \hat{b} на значение предиктора x_j для оцениваемого объекта:

$$Sb = \hat{b}_1 x_1 + \hat{b}_2 x_2 + \dots + \hat{b}_k x_k. \quad (32)$$

Последовательность действий по переводу коэффициентов модели в баллы — масштабированию переменной состоит в последовательном выполнении трех операций:

- определение диапазона числовой шкалы с минимальным и максимальным значением;
- определение количества баллов, удваивающее вероятность быть признанным «хорошим» заемщиком;
- определение значения шкалы, в которой достигается заданное отношение вероятностей плохих и хороших результатов.

Для более точной оценки скорингового балла формулу (32) можно дополнить несколькими параметрами:

$$Sb = A + R \hat{b}_j, \quad (33)$$

где A — смещение. Параметр R вычисляется по следующей формуле:

$$R = \frac{D}{\ln(2)}, \quad (34)$$

где D — баллы, удваивающие шансы стать «хорошим» объектом. Смещение можно вычислить по формуле:

$$A = B - R \ln(C), \quad (35)$$

где B — значение, при котором соотношение шансов составляет C . Множитель A был принят в расчет с положительным, либо отрицательным знаком в зависимости от логической составляющей предиктора. Например, с увеличением сокращения возраста транспортного средства наступление негативного события снижается, однако с увеличением совокупного пробега — все наоборот. Целью данного преобразования является перевод скоринговых баллов в логичную и понятную форму.

2. Результаты реализации и проверки скоринг-модели при наличии экспертных ограничений

2.1. Калибровка и оценка качества модели логистической регрессии

Проведем оценку качества скоринговых моделей, которые строились с учетом и без учета экспертных ограничений на нескольких выборках разного объема. Общая выборка включает 39418 записей о транспортных средствах, из которых 91,4 % — «хорошие», а 8,6 % — «плохие». Результаты расчетов и моделирования представлены в таблице 1.

Таблица 1

Результаты оценки качества моделей в зависимости от экспертных ограничений и объема выборки

Объем выборки	Сегмент юр. лиц при экспертных ограничениях	Сегмент юр. лиц без экспертных ограничений	Сегмент ИП с экспертными ограничениями	Сегмент ИП без экспертных ограничений
Общая выборка	0,8206	0,8177	0,8169	0,8188
80 % от общей выборки	0,7749	0,7689	0,7984	0,7752
20 % от общей выборки	0,7765	0,7092	0,7963	0,7593

На полных выборках методы ведут себя практически одинаково (рис. 1).

Проведем исследования, сократив объем исторической выборки до 80 и 20 % от общей выборки. Результаты экспериментов приведены на рисунках 2 соответственно.

Для понимания сути ошибок I и II рода рассмотрим таблицу сопряженности (см. табл. 2), которая строится на основе результатов классификации моделью и фактической принадлежностью примеров к классам ошибок. Зависимость верно классифицируемых объектов положительного класса от ложно положительно классифицируемых объектов негативного класса, иными словами, верно классифицированных «хороших» транспортных средств от неверно классифицированных «хороших» транспортных средств, которые по результатам проверки дали отрицательный производственный результат.

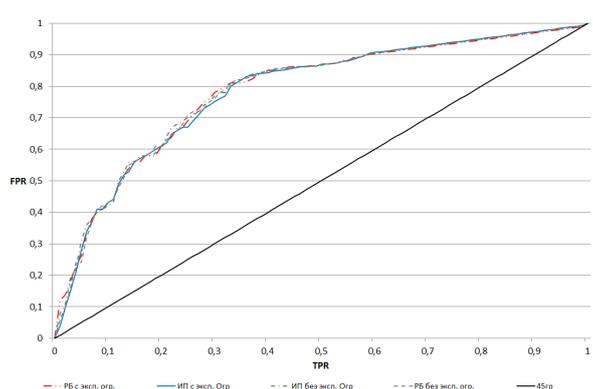


Рис. 1. Кривая скоринг-модели на общей выборке

TP (True Positives) — верно классифицированные «хорошие» транспортные средства, которые показали положительный производственный результат.

TN (True Negatives) — верно классифицированные отрицательные примеры.

FN (False Negatives) — «хорошие» транспортные средства, классифицированные как отрицательные (ошибка I рода). «Ложный пропуск» — когда интересующее нас событие ошибочно не обнаруживается.

FP (False Positives) — отрицательные примеры, классифицированные как положительные (ошибка II рода).

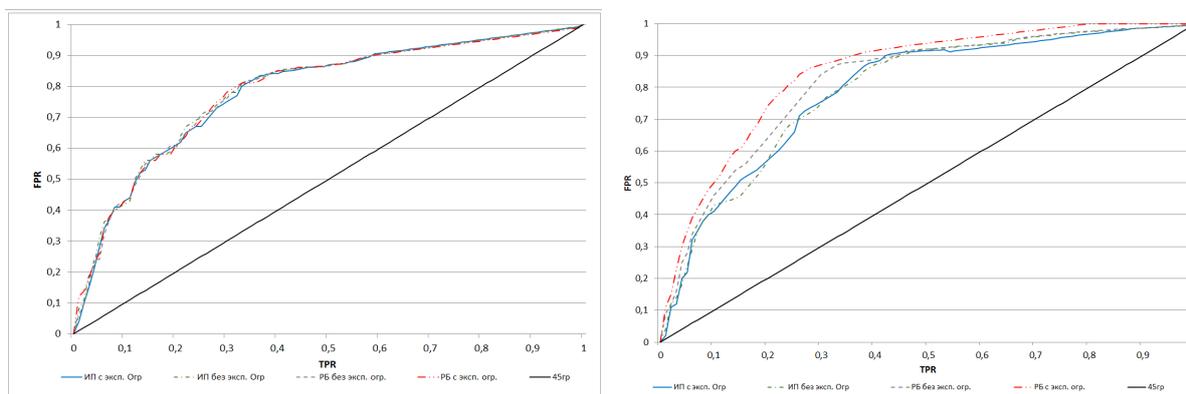


Рис. 2. Кривые скоринг-модели на 80 и 20 % от общей выборки

Таблица 2

Таблица сопряженности

Модель	Фактически положительно	Фактически отрицательно
Положительно	TP	FP
Отрицательно	FN	TN

Заключение

В результате выполнения данной работы удалось применить методы кредитного скоринга для фильтрации информационного потока, поступающего от дистрибьютеров, сервисных компаний и филиалов организации. Разработанная модель, реализованная в

виде программного продукта, позволила сократить на 76–79 % объем информации, требующей детального изучения. Качество фильтрации с использованием модели составило по ошибкам второго рода 2,7 %, а по ошибкам первого рода — 7,2 %. Применение данной модели как части информационной системы поддержки принятия управленческих решений при эксплуатации транспортных средств на природном газе позволило сократить время принятия решений и выбора альтернатив на 50 % и освободить от рутинной работы часть сотрудников аналитического отдела.

СПИСОК ЛИТЕРАТУРЫ

1. Абричкина, А. Г. Структура скоринг-модели анализа кредитных рисков / А. Г. Абричкина // ДМК Пресс. — 2003. — Т. 8. — С. 60–61.
2. Васильев, Ю. Н. Газозаправка транспорта / Ю. Н. Васильев, А. И. Гриценко, К. Ю. Чириков. — М. : Недра, 1995. — 446 с.
3. Вишняков, И. В. Методы и модели оценки кредитоспособности заемщиков / И. В. Вишняков. — СПб. : СПбГИЭА, 1998. — 267 с.
4. Дедков, В. К. Косвенные методы прогнозирования надежности / В. К. Дедков. — М. : ВЦ им. А.А. Дородницына, 2006. — 272 с.
5. Евстифеев, А. А. Основы логико-вероятностного анализа безопасности транспортных средств на газовом топливе / А. А. Евстифеев, С. В. Люгай. — М. : ООО «Газпром ВНИИГАЗ», 2017. — 206 с.
6. Евстифеев, А. А. Аналитическое моделирование безопасности и отказоустойчивости сложных технических систем / А. А. Евстифеев. — М. : НИЯУ «МИФИ», 2010. — 252 с.
7. Кочедыков, Д. А. Система кредитного скоринга на основе логических алгоритмов классификации / Д. А. Кочедыков // Математические методы распознавания образов. — 2005. — Т. 12. — С. 349–353.
8. Куртова, Л. Н. Основы математической логики / Л. Н. Куртова. — Белгород : Изд-во НИУ «БелГУ», 2018. — 85 с.
9. Льюис, Дж. Ядро Oracle. Внутреннее устройство для администраторов и разработчиков баз данных / Дж. Льюис. — М. : ДМК Пресс, 2015. — 372 с.
10. Маккинни, У. Python и анализ данных / У. Маккинни. — М. : ДМК Пресс, 2015. — 482 с.
11. Ноутон, П. Java2. В подлиннике. Наиболее полное руководство / П. Ноутон. — СПб. : БХВ-Петербург, 2007. — 1067 с.
12. Открытый курс машинного обучения: анализ временных рядов с помощью Python. — Электрон. текстовые дан. — Режим доступа: <https://habr.com/ru/company/ods/blog/327242/>. — Загл. с экрана.
13. Решение проблемы оптимального синтеза технологических процессов сложных систем / Е. Н. Малыгин, В. А. Немтинов, Ж. Е. Зимнухова, Ю. В. Немтинова // Вестн. Тамбов. ун-та. Серия: Естественные и технические науки. — 2002. — Т. 7, № 2. — С. 242–245.
14. Рашка, С. Python и машинное обучение / С. Рашка. — М. : ДМК Пресс, 2017. — 418 с.
15. Рыбина, Г. В. Применение интеллектуального анализа данных для построения баз знаний интегрированных экспертных систем / Г. В. Рыбина // Авиакосмическое приборостроение. — 2012. — Т. 11. — С. 36–53.
16. Сравнение экономических показателей при использовании жидкого моторного и газомоторного топлив / С. В. Люгай, А. А. Евстифеев, В. В. Тимофеев, М. Л. Балашов, Ю. Н. Дрыгина // Транспорт на альтернативном топливе. — 2013. — № 5 (35). — С. 14–19.
17. Шестакова, Е. С. Анкетный опрос / Е. С. Шестакова, М. И. Кулик, А. Ю. Скриган // Метод кейсов в комплексных социально-экологических исследованиях. — Псков : Изд-во Псков. гос. ун-та, 2017. — С. 77–88.

18. Fan, J. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties / J. Fan, R. Li // *Journal of the American Statistical Association*. — 1996. — Vol. 6. — P. 1348–1360.

REFERENCES

1. Abrichkina A.G. *Struktura skoring-modeli analiza kreditnykh riskov* [Scoring Structure — Credit Risk Analysis Models]. *DMK Press*, 2003, vol. 8, pp. 60-61.
2. Vasilyev Yu.N., Gritsenko A.I., Chirikov K.Yu. *Gazozapravka transporta* [Gas Filling of Transport]. Moscow, Nedra Publ., 1995. 446 p.
3. Vishnyakov I.V. *Metody i modeli otsenki kreditosposobnosti zaemshchikov* [Methods and Models for Assessing the Creditworthiness of Borrowers]. Saint Petersburg, SPbGIEA, 1998. 267 p.
4. Dedkov V.K. *Kosvennye metody prognozirovaniya nadezhnosti* [Indirect Methods for Predicting Reliability]. Moscow, VTs im. A.A. Dorodnitsyna, 2006. 272 p.
5. Evstifeev A.A., Lyugay S.V. *Osnovy logiko-veroyatnostnogo analiza bezopasnosti transportnykh sredstv na gazovom toplive* [Fundamentals of Logical-Probabilistic Safety Analysis of Gas-Fueled Vehicles]. Moscow, OOO “Gazprom VNIIGAZ”, 2017. 206 p.
6. Evstifeev A.A. *Analiticheskoe modelirovaniye bezopasnosti i otkazoustoychivosti slozhnykh tekhnicheskikh sistem* [Analytical Modeling of Security and Fault Tolerance of Complex Technical Systems]. Moscow, NIYaU “MIFI”, 2010. 252 p.
7. Kochedykov D.A. *Sistema kreditnogo skoringa na osnove logicheskikh algoritmov klassifikatsii* [Credit Scoring System Based on Logical Classification Algorithms]. *Matematicheskie metody raspoznavaniya obrazov*, 2005, vol. 12, pp. 349-353.
8. Kurtova L.N. *Osnovy matematicheskoy logiki* [Foundations of Mathematical Logic]. Belgorod, Izd-vo NIU BelGU, 2018. 85 p.
9. Lyuis Dzh. *Yadro Oracle. Vnutrennee ustroystvo dlya administratorov i razrabotchikov baz dannykh* [The Core of Oracle. Backend for Administrators and Database Developers]. Moscow, DMK Press Publ., 2015. 372 p.
10. McKinney W. *Python i analiz dannykh* [Python and Data Analysis]. Moscow, DMK Press Publ., 2015. 482 p.
11. Nouton P. *Java2. V podlinnike. Naibolee polnoe rukovodstvo* [Java2. In the Original. The Most Complete Guide]. Saint Petersburg, BKhV-Peterburg Publ., 2007. 1067 p.
12. *Otkrytyy kurs mashinnogo obucheniya: analiz vremennykh ryadov s pomoshchyu Python* [Open Machine Learning Course: Time Series Analysis with Python]. URL: <https://habr.com/ru/company/ods/blog/327242/>.
13. Malygin E.N., Nemtinov V.A., Zimnukhova Zh.E., Nemtinova Yu.V. *Reshenie problemy optimalnogo sinteza tekhnologicheskikh protsessov slozhnykh sistem* [Solving the Problem of Optimal Synthesis of Technological Processes of Complex Systems]. *Vestn. Tambov. un-ta. Seriya: Estestvennye i tekhnicheskie nauki*, 2002, vol. 7, no. 2, pp. 242-245.
14. Rashka S. *Python i mashinnoe obuchenie* [Python and Machine Learning]. Moscow, DMK Press Publ., 2017. 418 p.
15. Rybina G.V. *Primeneniye intellektualnogo analiza dannykh dlya postroeniya baz znaniy integrirovannykh ekspertnykh sistem* [Application of Data Mining for Building Knowledge Bases of Integrated Expert Systems]. *Aviakosmicheskoe priborostroeniye*, 2012, vol. 11, pp. 36-53.
16. Lyugay S.V., Evstifeev A.A., Timofeev V.V., Balashov M.L., Drygina Yu.N. *Sravneniye ekonomicheskikh pokazateley pri ispolzovanii zhidkogo motornogo i gazomotornogo topliv* [Comparison of Economic Indicators When Using Liquid Motor and Gas Motor Fuels]. *Transport na alternativnom toplive*, 2013, no. 5 (35), pp. 14-19.
17. Shestakova E.S., Kulik M.I., Skrigan A.Yu. *Anketnyy opros* [Questionnaire Survey]. *Metod keysov v kompleksnykh sotsialno-ekologicheskikh issledovaniyakh*. Pskov, Izd-vo Pskov. gos. un-ta, 2017, pp. 77-88.
18. Fan J., Li R. Variable Selection Via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 1996, vol. 6, pp. 1348-1360.

**FORECASTING THE RISKS OF AN ORGANIZATION
OPERATING NATURAL GAS VEHICLES
USING A SCORING MODEL OF LOGISTIC REGRESSION
IN THE PRESENCE OF EXPERT RESTRICTIONS**

Andrey A. Evstifeev

Candidate of Technical Sciences, Associate Professor,
Department 12 of Computer Systems and Technologies,
National Research Nuclear University,
Moscow Engineering Physics Institute
aaevstifeev@mephi.ru
<https://orcid.org/0000-0003-3354-0487>
Kashirskoe Shosse, 31, 115409 Moscow, Russian Federation

Abstract. The paper proposes a method and describes a mathematical model for express analysis of the attractiveness of the operation of vehicles running on natural gas for a motor transport company. The proposed solution is based on a logistic regression scoring model used by banks to assess the creditworthiness of a borrower. To improve the quality of the results, the model is extended with a set of expert restrictions formulated in the form of rules. During the analysis, signs were identified that require quantization, since individual intervals of values turned out to be associated with risk in different ways. The developed mathematical model is implemented in the form of software in a high-level programming language, the information of the model is stored in a database management system and is integrated with an information system for supporting management decisions when operating vehicles on natural gas. The developed mathematical model was tested on a test training sample. The test results showed a satisfactory accuracy of the proposed model at the level of 77% without the use of expert restrictions and 79% with their use. At the same time, the share of Type II errors was 2.7%, and Type I errors were 7.2%, which indicates that the model is quite conservative, and a relatively high proportion of vehicles that meet the requirements were rejected.

Key words: scoring model, logistic regression, vehicle operation, decision making, complex technical system.